NTP
National Toxicology Program

Bioinformatics-based identification of assays that inform on disease hazard

Scott S. Auerbach, Ph.D., DABT
National Institute of Environmental Health Sciences

NTP Board of Scientific Counselors Meeting
December 1, 2010

Outline

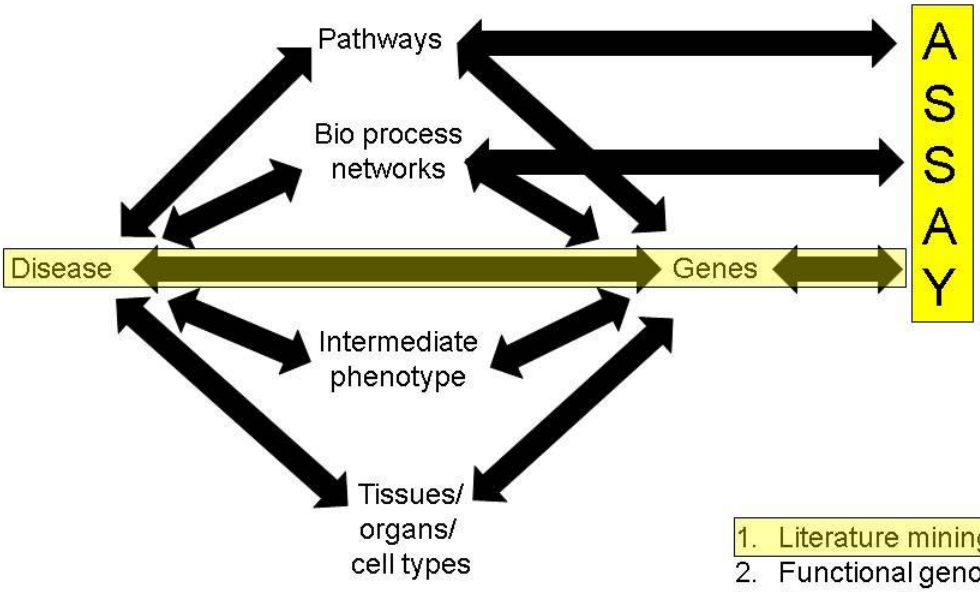
- Goals, Overall Concept and Approach
- Date Resources and Critical Concepts
- Results of Initial Analysis
- Future Work

Goals (fill in variables):

1. Disease Y hazard is queried most effectively by assay A,B,C….

2. Assay X queries disease hazard A, B, C…

From diseases to genes to assays and back again



1. Literature mining
2. Functional genomics
3. Genetics/Genomics

Approach

- Assemble list of all genes in the human genome and associated annotations
- Extract Disease-Gene relationships from database resources
- Map associated genes to human gene reference list
- Calculate cumulative score for all genes for a given disease
- Determine assay development tractability/feasibility based on a "druggability" score

Data Resources: Literature mining

- Comparative Toxicogenomics Database (CTD), Ingenuity, GeneGo, OMIM
  - Hand curated associations between, diseases, genes, pathways, biological processes
- CoPub
  - Automated curation of associations between diseases, genes, tissues, pathways, biological processes, and pathways
  - Derived from text mining of abstracts
- GeneCards and Entrez gene
  - Automated curation of diseases/gene associations
  - Derived from text mining of abstracts
- Phenopedia (HuGE Navigator)
  - Automated curation of diseases/gene associations
  - Derived from text mining of abstracts
  - Focus on genetic association studies

Data Resources: Functional genomics

- NextBio
  - Contains genomic signatures from a large faction of public data deposited in GEO, ArrayExpress, dbGaP, and Cosmic databases
    - RNA expression, CNV, Somatic mutation, DNA methylation, Histone modification, DNA protein interaction, etc.
  - Identifies disease/gene/pathway relationships based on differential gene expression, DNA methylation, etc.
- Unigene Body Atlas
  - Identifies the relationship between tissues and gene-based patterns of tissue-specific expression
- GeneoGo and Ingenuity
  - Identifies the relationship between pathways/biological processes that are enriched in the gene expression from prediseased and diseased tissue

Data Resources: Genetics

- NextBio
  - Identifies relationships between diseases/intermediate phenotypes and genes based on OMIM entries and meta-analysis of GWAS studies deposited in dbGaP
- Phenopedia
  - GWAS data curation

Critical Concept: Druggability

- The protein encoded by a gene is considered druggable if it contains a conserved protein domain that has been shown to bind to small molecules
  - Examples: Nuclear receptors, G-protein coupled receptors, Kinases
- Druggable genome sources:
  - Ensembl
  - DrugBank
  - InterPro-BLAST
  - BioLT
  - Quiagen
  - Dharmacon
- There are 6 druggable genome resources, hence a gene gets a druggability score from 0-6

Type 1 Diabetes Prioritized Assay Targets

	Hand curated				Automated curation						
Official Gene Name	CTD	Ingenuity	GeneGo	OMIM	Entrez Gene	CoPub	GeneCards	Phenopedia	Sum	Druggability Score	Existing Assays
cytotoxic T-lymphocyte-associated protein 4	2	2	2	2	1	1	1	1	12	4	0
insulin	2	2	2	2	1	1	1	1	12	3	0
SMT3 suppressor of mif two 3 homolog 4 (S. cerevisiae)	2	2	2	2	1	1	1	1	12	0	0
protein tyrosine phosphatase, non-receptor type 22 (lymphoid)	2	2	2	2	1	0	1	1	11	5	0
interleukin 2 receptor, alpha	2	2	2	2	1	0	1	1	11	4	0
inositol 1,4,5-triphosphate receptor, type 3	2	2	2	2	1	0	1	1	11	3	0
major histocompatibility complex, class II, DQ beta 1	2	2	2	2	1	0	1	1	11	2	0
forkhead box P3	2	2	2	2	1	0	1	1	11	2	0
SH2B adaptor protein 3	2	2	2	2	1	0	1	1	11	0	0
interferon induced with helicase C domain 1	2	2	2	2	1	0	1	1	11	0	0

*T1D hazard may be queried most effectively by assays that evaluate CTLA4, PTPN22, IL2 activity

Druggable targets that are not related to T1D biology

ATP-binding cassette, sub-family A (ABC1), member 1	0	0	0	0	0	0	0	0	0	6	0
dopamine receptor D2	0	0	0	0	0	0	0	0	0	6	1
nuclear receptor subfamily 5, group A, member 1	0	0	0	0	0	0	0	0	0	6	1

CTLA4 polymorphisms are associated with autoimmune disease, Type 1 Diabetes

Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease

Hironori Ueda[1], Joanna M. M. Howson[1]*, Laura Esposito[1]*, Joanne Heward[2]*, Hywel Snook[1], Giselle Chamberlain[1], Daniel B. Rainbow[1], Kara M. D. Hunter[1], Annabel N. Smith[1], Gianfranco Di Genova[1]†, Mathias H. Herr[1]†, Ingrid Dahlman[1]†, Felicity Payne[1], Deborah Smyth[1], Christopher Lowe[1], Rebecca C. J. Twells[1], Sarah Howlett[1], Barry Healy[1], Sarah Nutland[1], Helen E. Rance[1], Vin Everett[1], Luc J. Smink[1], Alex C. Lam[1], Heather J. Cordell[1], Neil M. Walker[1], Cristina Bordin[1]†, John Hulme[1], Costantino Motzo[3], Francesco Cucca[3], J. Fred Hess[4], Michael L. Metzker[4]†, Jane Rogers[5], Simon Gregory[5], Amit Allahabadia[2]†, Ratnasingam Nithiyananthan[2], Eva Tuomilehto-Wolf[6], Jaakko Tuomilehto[6,7], Polly Bingley[8], Kathleen M. Gillespie[8], Dag E. Undlien[9]†, Kjersti S. Rønningen[10], Cristian Guja[11], Constantin Ionescu-Tirgovişte[11], David A. Savage[12], A. Peter Maxwell[13], Dennis J. Carson[14], Chris C. Patterson[15], Jayne A. Franklyn[2], David G. Clayton[1], Laurence B. Peterson[16], Linda S. Wicker[1], John A. Todd[1] & Stephen C. L. Gough[2]

Nature. 2003 May 29;423(6939):506-11

Type 2 Diabetes Prioritized Assay Targets

	Hand curated				Automated curation						
Official Gene Name	**CTD**	**Ingenuity**	**GeneGo**	**OMIM**	**Entrez Gene**	**CoPub**	**GeneCards**	**Phenopedia**	**Sum**	**Druggability Score**	**Existing Assays**
peroxisome proliferator-activated receptor gamma	2	2	2	2	1	1	1	1	12	6	1
potassium inwardly-rectifying channel, subfamily J, member 11	2	2	2	2	1	1	1	1	12	6	1
hepatocyte nuclear factor 4, alpha	2	2	2	2	1	1	1	1	12	6	1
ATP-binding cassette, sub-family C (CFTR/MRP), member 8	2	2	2	2	1	1	1	1	12	6	0
ectonucleotide pyrophosphatase/phosphodiesterase 1	2	2	2	2	1	1	1	1	12	5	0
glucagon receptor	2	2	2	2	1	1	1	1	12	4	0
neurogenic differentiation 1	2	2	2	2	1	1	1	1	12	3	0
lipase, hepatic	2	2	2	2	1	1	1	1	12	3	0
insulin receptor substrate 2	2	2	2	2	1	1	1	1	12	3	0
insulin receptor substrate 1	2	2	2	2	1	1	1	1	12	3	0

PPARG polymorphisms lead to differential susceptibility to T2D

The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes

David Altshuler[1,2,3*], Joel N. Hirschhorn[1,3,4*], Mia Klannemark[5], Cecilia M. Lindgren[1,5], Marie-Claude Vohl[6], James Nemesh[1], Charles R. Lane[1], Stephen F. Schaffner[1], Stacey Bolk[1], Carl Brewer[6], Tiinamaija Tuomi[5,7], Daniel Gaudet[8], Thomas J. Hudson[1,6], Mark Daly[1], Leif Groop[5] & Eric S. Lander[1,9]

Nat Genet. 2000 Sep;26(1):76-80

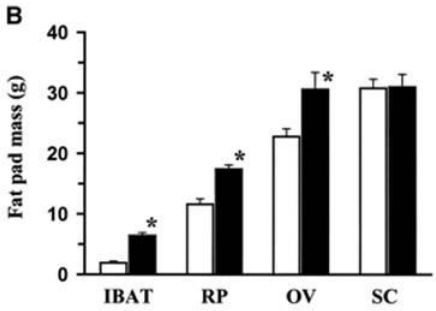
Obesity Prioritized Assay Targets

	Hand curated				Automated curation						
Official Gene Name	CTD	Ingenuity	GeneGo	OMIM	Entrez Gene	CoPub	GeneCards	Phenopedia	Sum	Druggability Score	Existing Assays
peroxisome proliferator-activated receptor gamma	2	2	2	2	1	1	1	1	12	6	1
peroxisome proliferator-activated receptor alpha	2	2	2	2	1	1	1	1	12	6	1
serpin peptidase inhibitor, clade E (nexin, plasminogen activator	2	2	2	2	1	1	1	1	12	5	1
peroxisome proliferator-activated receptor delta	2	2	2	2	1	1	1	1	12	5	1
insulin receptor	2	2	2	2	1	1	1	1	12	5	1
adrenergic, beta-3-, receptor	2	2	2	2	1	1	1	1	12	5	1
proprotein convertase subtilisin/kexin type 1	2	2	2	2	1	1	1	1	12	5	0
ectonucleotide pyrophosphatase/phosphodiesterase 1	2	2	2	2	1	1	1	1	12	5	0
cannabinoid receptor 1 (brain)	2	2	2	2	1	1	1	1	12	5	0
neuropeptide Y	2	2	2	2	1	1	1	1	12	4	1

PPARγ regulates adiposity

Effects of Pioglitazone on Adipose Tissue Remodeling Within the Setting of Obesity and Insulin Resistance

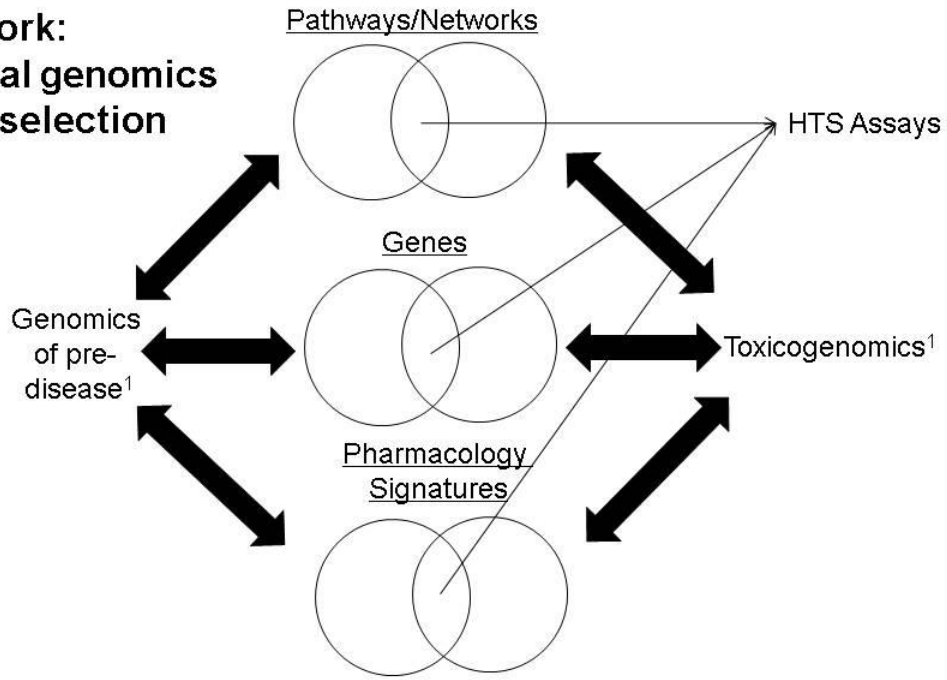
Christopher J. de Souza,[1] Michele Eckhardt,[1] Karen Gagen,[1] Mei Dong,[1] Wei Chen,[1] Didier Laurent,[2] and Bryan F. Burkey[1]



*27 days of pioglitazone (PPARY agonist) leads to increased fat pad mass

Diabetes. 2001 Aug;50(8):1863-71.

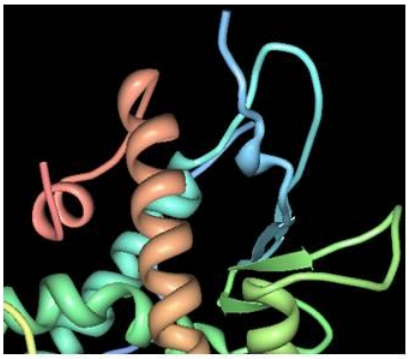
Future work: Functional genomics in assay selection



[1] Data to evaluate these relationships comes from NextBio, CEBS, and DrugMatrix

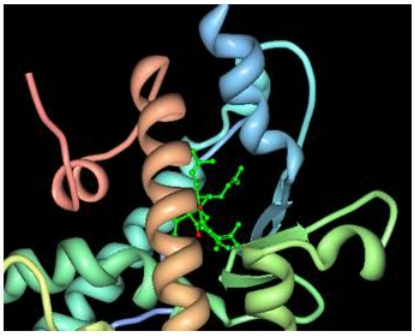
Future Work: Consideration of target promiscuity

Human PXR Ligand Binding Domain

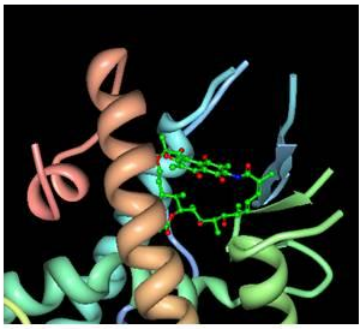


Apo

Hyperforin

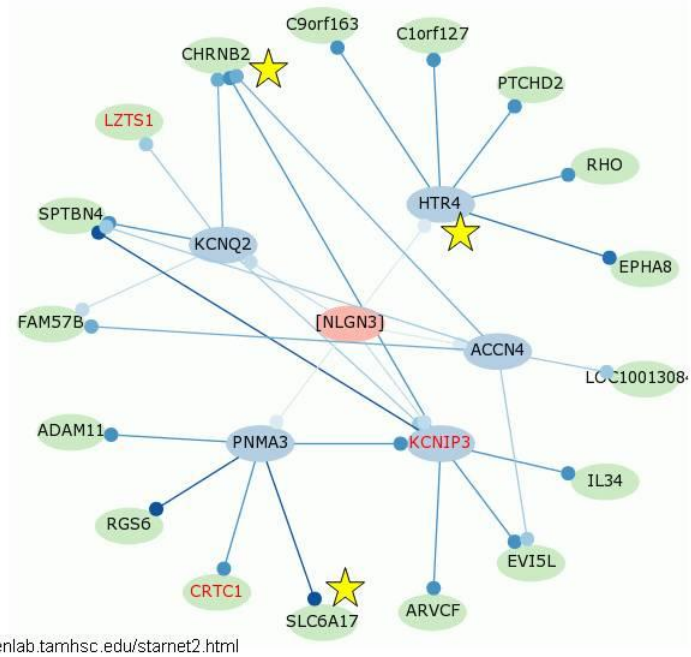


Rifampicin



Data derived from publications from the Redinbo Lab (UNC)

Future Work: Integrate gene-gene relationships to identify novel targets



StarNet2: http://vanburenlab.tamhsc.edu/starnet2.html

Summary

- We are integrating multiple genomics/bioinformatics resources to identify genes/pathways associated with disease
- Based on the associations, we plan to identify or develop *in vitro* assays that will predict chemical hazard in a disease-centric fashion
- We anticipate that the findings from screening strategies base don disease-centric assays will facilitate both the prioritization of chemicals for focused *in vivo* studies (i.e., targeted testing) and the development of integrated testing strategies that will be human centric while being more cost effective, more efficient, and more informative
- Independent efforts in this area of research are ongoing at NCGC, EPA and have been published by a group at NIEHS
  - A comparison of results will be performed following completion of these efforts

NTP Workshop:
Role of Environmental Chemicals in the Development of Diabetes and Obesity

January 11-13, 2011

Raleigh Marriott Crabtree Valley • 4500 Marriott Drive

http://cerhr.niehs.nih.gov/evals/diabetesobesity/index.html

Acknowledgements

- NTP/CERHR
  - Kristine Thayer
  - Vickie Walker
- EPA
  - Richard Judson
- NCGC
  - Allison Mandich